DOCUMENT RESUME

ED 148 887 95 TM 006 780

AUTHOR Crawford, John
TITLE Relationships Between Grade Point Average, Response Mode, Difficulty Level, and Learning. Report No. 75-10.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO UTR&D-R-75-10
PUB DATE 75
CONTRACT NIE-C-74-0089; OEC-6-10-108
NOTE 31p.; Revision of a paper presented at the Annual Meeting of the Southwestern Psychological Association (Houston, Texas, April, 1975)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Academic Ability; Academic Achievement; College Students; *Complexity Level; Grade Point Average; Higher Education; Instructional Design; Instructional Materials; *Learning Activities; Learning Characteristics; Learning Theories; Motivation; *Performance Factors; Pretests; *Programed Materials; Psychophysiology; Questioning Techniques; *Response Mode; Student Role; Teaching Techniques
IDENTIFIERS Errorless Learning

ABSTRACT

The effects of response mode and complexity level of instructional materials, and the effects of student ability (as indicated by grade point average) on learning performance were studied. Four programmed instructional units on an unfamiliar subject, physiological psychology, were presented to 159 college students. Three of the units required the learners to construct fill-in-the-blank responses to questions of varying complexity; the programmed materials were categorized according to low, medium, or high error rate. The fourth set of materials required reading only. Learning was assessed with a 30-item multiple choice pretest and post test. High ability subjects performed better than low ability subjects, regardless of treatment. Curve-fitting regression techniques showed that different ability groups have different learning curves when plotted against the error rate of the materials. These trends are consonant with achievement motivation theory predictions. The reading-only mode was found to be superior to the constructed response; this result may be explained by the familiarizing effect of the pretest. Related studies are briefly discussed. (GDC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED148887

4021

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Research and Development Center for Teacher Education

UTR&D

The University of Texas at Austin

TM006 780

The main objective of the CORRELATES OF EFFECTIVE TEACHING PROJECT is to expand the number of teaching principles based on documented findings from systematic classroom research. The principles and procedures [illegible] on the basis of [illegible] with teachers and school personnel. Emphasis is on the study of the classroom to discover how these principles can be conducted to the greatest advantage of teachers and individual students.

One of the project's major efforts was a two-year study of teaching effectiveness involving the examination of the classroom behavior of teachers consistent in producing student learning gains.

Since 1974 three other major data collection efforts were initiated and completed.

(1) STUDENT ATTRIBUTE STUDY which looked at student characteristics and behaviors and their effects on teachers.

(2) FIRST GRADE READING GROUP STUDY, an experimental study designed to test the effectiveness of selected group management techniques in teaching reading.

(3) JUNIOR HIGH SCHOOL STUDY, a follow up on earlier work from the second and third grades of the influence of teacher characteristics and behavior on students' cognitive and affective gains.

RELATIONSHIPS BETWEEN GRADE POINT AVERAGE, RESPONSE MODE, DIFFICULTY LEVEL, AND LEARNING[1]

John Crawford

Research and Development Center for Teacher Education

The University of Texas at Austin

Report No. 75-10

[1] An earlier version of this paper was presented at the Southwestern Psychological Association's annual meeting, Houston, Texas, April, 1975.

This study was supported in part by the National Institute of Education Contract OEC 6-10-108, Research and Development Center for Teacher Education, and by Contract NIE-C-74-0089, Correlates of Effective Teaching. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

Abstract

Using programmed instruction format materials, the effects of response mode, difficulty level, and college GPA were studied, using 159 college students as subjects. Regardless of treatment, high GPA subjects did better than low. The reading-only mode was superior to constructed response. Curve-fitting regression techniques yielded evidence that different GPA groups have different learning curves when plotted against number of errors on the teaching materials. The superiority of the reading-only materials fits a "familiarity" interpretation -- the pretest served as a familiarizing agent. The differentially curvilinear trends for high and low GPA groups are consonant with achievement motivation theory predictions. Discussions of factors such as time on task, learner participation, and matching the instruction to current cognitive structures are also presented.

Relationships between Grade Point Average, Response Mode, Difficulty Level, and Learning

The present study was designed to test assumptions concerning optimal instructional design using programmed instruction (PI) format teaching materials. The variables of interest included: difficulty level of materials, response mode, and subject ability as measured by overall college GPA. Learning was assessed for each treatment group with a multiple choice post-test covering the same subject matter as the teaching materials.

Investigations of PI materials and their relation to learning gains have the potential both for specifying the types of PI materials most effective in producing learning, and, in a broader sense, for serving as tools for testing hypotheses concerning various instructional variables. To date, these efforts have generally been productive, although not always conclusive.

Since the popularization of teaching machines and PI, there have been advocates of errorless (or near-errorless) learning (Skinner, 1961; Pysh, Blank, and Lambert, 1969; Holland, 1967) and critics of this concept (Pressey and Kinzer, 1964; Crowder, 1962). While there has been no definite resolution of the conflict, it is now a widely accepted practice to employ materials with low, near-zero error rates. The major rationale for this strategy is that if subjects proceed through materials in very small steps (with few errors), the reinforcement value of their high success rates will allow for maximum assimilation of information.

Process-product data from the Texas Teacher Effectiveness Project (Brophy and Evertson, in press) suggested non-linear relationships between teaching effectiveness measures and success rates of second and third grade students in public academic response opportunities. The optimal rates of correct answers to teachers' questions were significantly less than 100%, ranging from 75-85%. The optimal level of correct responses varied with school SES; the error rate at which the low SES students learned best was lower than the error rate that allowed the high SES students to make the largest gains. That is, low SES students did better with easier questions and high SES students learned most when challenged with more difficult questions. Thus, while PI materials generally have near-zero error rates, teachers seem to question students optimally with error rates around 15-25%.

The issue of which response mode is most effective has been studied by using PI in different forms (constructed response, multiple choice, reading-only). Subjects are usually given instruction in different modes covering the same material; then, posttest scores are examined for any differences in learning between the groups.

Using the Holland and Skinner (1961) operant theory program, Williams (1963) found that reading materials were more efficient (number of correct responses per unit of time) than constructed response or multiple choice type instruction. However, for total "amount learned" (number of correct responses), the constructed response and multiple choice groups were significantly superior to the reading mode groups, with or without emphasis. There were no differences between either the constructed response and multiple choice

groups, or the reading groups, with or without underlining. There was a significant main effect for aptitude: subjects who scored high on the SAT did better than the ones who had low SAT scores, regardless of mode of presentation. Goldbeck and Campbell (1962) also demonstrated that the most efficient type of teaching materials is reading mode materials; i.e., no responses are required of the learner.

Using technical and non-technical versions of the myocardial infarction program, Tobias (1969) reported a significant interaction between response mode and familiarity. For technical, unfamiliar subject matter, constructed response materials were superior to reading mode materials. However, when the subject matter was familiar, the reading mode materials were more effective than the constructed response materials. (Planned t-test comparisons, conducted by the present author on Tobias' data, indicated that both the low and high creative groups evidenced non-significant trends of reading mode materials producing more learning than the constructed response materials when the subject matter was familiar: t=1.23, df=48, $.15 > p > .10$ for the low creative group, and t=.80, df=48, $.25 > p > .20$ for the high creative group.) There were no significant interactions between creativity and response mode or between creativity and subject matter.

Abramson and Kagen (1975) also hypothesized that the type of response mode which is most effective depends upon the familiarity of the material. They manipulated familiarity by presenting half their subjects with lists of relevant technical terms prior to the programmed instruction. The subjects were assigned either to reading mode or constructed response groups. The

programs used were versions of the myocardial infarction program used in other studies. Upon completion of the PI materials, subjects were given immediate posttests, and then delayed posttests one week later. The posttests covered both verbal and electrocardiogram tracing material.

Their main effects results were inconclusive: familiarization groups had higher scores than nonfamiliarization groups, and constructed response was better than reading mode, but only for one of four criteria measures (delayed posttest tracing score, in each case). Males did significantly better than females on the posttest tracing score. However, interaction hypotheses were supported on three measures, and approached significance on the fourth. For unfamiliar material, constructed response was better than reading. However, for familiar material, reading was better than constructed response on two of the four criteria: posttest tracing (t=1.71, df=27, $p<.05$) and delayed posttest tracing (t=2.22, df=27, $p<.025$). Graphic data are also presented which show that reading-only instruction was better than constructed response for familiar verbal material -- on both the posttest and delayed posttest -- for males, though not for females.

Tobias (1973) reviewed the issue of the effectiveness of different response modes, presenting some data indicating that constructed response produced more learning than reading, along with some other studies that found no significant differences between constructed response and reading-only. He discussed and rejected the blackout ratio factor as incapable of accounting for the constructed response versus reading-only results. Tobias hypothesized that "time on task" may be an important underlying factor in the studies which find constructed response more effective than reading materials. He also sug-

gested generalizing the PI research results on these instructional variables by using materials other than the typically used Holland and Skinner (1961) and heart disease diagnosis programs.

The present study was designed for the purposes of:

1.) First, testing for an overall treatment effect; do subjects learn different amounts according to the set of materials used?

2.) Analyzing more specific treatment effects: low versus medium versus high error rate materials and reading-only versus constructed response materials.

3.) Checking for aptitude-treatment interactions; low and high GPA groups may learn optimally with different sets of materials.

Method

A standard covariance factorial design was employed. The treatments consisted of the different sets of teaching materials. The independent variables were: difficulty level of the materials, response mode (reading versus constructed response), and subject GPA level. The critical measure was the subject's score on the multiple choice posttest. The pretest scores were used as the adjusting covariable in Ward and Jennings (1973) linear models covariance procedure.

Subjects

One hundred and sixty-five volunteer subjects from a required introductory government course at the University of Texas completed their packets of teaching materials and tests. Data from six subjects were not used because

of failure to follow directions correctly. Eleven others failed to record their GPA's. So, the N for the first analysis of treatment effects was 159, the N for the difficulty level analyses was 109 (no reading mode subjects included), and the N for the response mode analyses was 148 (only excluding subjects who didn't record their GPA's). Subjects were randomly assigned to treatment groups: packets of materials were assembled in random order before they were passed out to the subjects.

Materials

A 30-item, multiple choice pretest-posttest was used to assess the amount of learning engendered by the teaching materials. This test covered the same subject matter as the materials. Test-retest correlations computed within each of the four treatment groups averaged .47. This figure indicates a significant relationship, but it also suggests that a large percentage of the posttest score variance was accounted for by the treatments. The content validity of the test rests on the fact that the information tested for, as well as the content of the items in the teaching materials, was derived from a standard college physiological psychology textbook (Morgan, 1965). The pretest scores did not differ significantly from what would be expected from chance guessing ($z = 1.33$, $p > .05$). Also, there were no between group differences in pretest scores: $F = .54$, $df = 3,155$, $p = .66$. This indicates that the randomization procedure was successful.

The four sets of PI format teaching materials were the other instruments used. These materials were designed specifically for use in this investigation. Three of the four sets were constructed response mode (fill in the blank). These different sets covered equal amounts of subject matter, but

differed in difficulty level (error rate). The fourth group used reading-only materials. This set was identical to the low error rate set except that the blanks were already filled in; the responses required in the other materials were typed in and underlined. These four sets of teaching materials constituted the experimental treatments.

Pilot testing had established the average rates of the constructed response materials at about 4% for the low error rate group, about 15% for the medium error rate group, and 42% for the high error rate group. The rates computed in this study were 3.75%, 18.75%, and 35.25%, respectively, for the low, medium, and high error rate groups.

The topic was physiological psychology, chosen primarily because of the lack of exposure of most subjects to this area. The questions were from four content areas: the central nervous system, neuronal physiology, the internal environment, and emotion. There were 10 frames for each of the four areas, with the answer to one frame given in the margin next to the subsequent frame on the following page. The backs of the sheets had a strip of paper taped to them to keep the subjects from seeing the answers through the sheet. This format allowed the subject to fill in the blank on one frame, turn to the next page, check his or her answer for correctness, and then proceed to the next frame.

Attention was given to construction procedures described in O'Day, Kulhavy, Anderson, and Malczynski (1971) in designing the materials. Difficulty level was manipulated by leaving out letters in partial presentations of words, by omitting "clueing phrases" which aid the subject in constructing the response, and/or by slightly changing the wording of the frames so as to make the answers less obvious. The required responses in parallel frames

in different sets of materials were always the same.

Procedure

The experimental procedure first involved giving verbal and written instructions to the subjects (including informing them that "this is a study trying to determine what types of teaching materials help students learn the most") and having them take the 30-item pretest. As soon as the subjects finished the pretest, they proceeded throught their set of materials. This was followed by the posttest, and then the subjects turned in their packets of tests and teaching materials. The experiment took place in a group testing situation, so monitors were used to keep improper use of materials to a minimum. Recording each subject's completion time was not feasible, since the subjects completed their materials in a group and turned in their packets when completed. However, it was clear that the reading-only subjects generally finished before the constructed response subjects. The amount of experimental time involved, including pretest, completion of materials, and posttest, was approximately 60 minutes. When all subjects had finished, the experimenters described the purpose of the study, possible expected results, and offered to give further feedback when data analyses were completed.

There were 39 subjects in the reading mode group, 39 in the low error rate group, 45 in the medium error rate group, and 36 in the high error rate group. Median splits were made within each of the treatment groups, allowing

for a high-low GPA distinction in each. The GPA distribution was approximately normal, ranging from 1.5 to 4.0, with both the mean and median at 3.0. While the cell N's were not equal, the "unequalness" should be random. The data analysis procedure did make appropriate adjustments for these unequal cell N's.

Results

The dependent measure was the number of correct responses on the multiple choice posttest. The distribution of these scores was skewed due to a posttest "ceiling" effect. In order to correct for this skewness, the normalizing transformation $\log (x + 1)$ was applied to the subjects' posttest scores. Winer (1971) describes this transformation as appropriate for normalizing skewed data. The subjects' pretest scores (number of items answered correctly) was the covariable in the linear models covariance analyses.

The first analysis, testing for an overall treatment effect, was a one-way, four level analysis of covariance. We see from Table 1 that the significant effect ($F=3.17$, $df=3,154$, $p<.03$) was due to a linear decrease in the dependent variable: the reading mode group scored the highest, followed by the low, then the medium, and then the high error rate groups.

Insert Table 1 About Here

The next set of analyses excluded the reading mode group and tested for a difficulty level main effect, a GPA main effect, and a difficulty level by GPA interaction. Median split, high-low groupings were used for the GPA

effects tests. The 2X3 table of means and standard deviations and the source data for these analyses are presented in Table 2.

Insert Table 2 About Here

Table 2 shows a highly significant main effect for GPA (F=13.09, df=1,102, $p < .001$), with the high GPA subjects scoring higher than the low GPA subjects, regardless of treatment. The linear decreasing trend found in the first analysis was not significant for the constructed response groups, though the low difficulty level groups (low and high GPA) were superior to the medium, which were higher than the high difficulty level groups. There was no significant difficulty level by GPA interaction.

Difficulty level apparently did not affect learning signficantly, at least when subjects were grouped by materials averaging 3.75%, 18.75%, and 35.25% errors. While the three constructed response sets of materials did average 3.75%, 18.75%, and 35.25% errors, there was considerable variation in the actual number of errors made by different subjects using identical sets of teaching materials. That is, some subjects using the low difficulty level materials made as many errors as other subjects assigned to the medium difficulty level group, and some subjects in the medium difficulty level group made as many errors as the high difficulty level group.

Given this overlap between groups whose average difficulty level differed considerably, it seemed reasonable to also analyze the posttest scores according to the actual number of errors the subjects made in responding to the

materials -- as opposed to the group (easy versus moderate versus difficult) to which the subjects were assigned. This analysis, which yielded results different from those reported in Table 2, was accomplished by curve-fitting regression techniques: the transformed posttest scores were regressed with the number of errors made on the constructed response teaching materials. The regression lines for the low and high GPA groups are plotted in Figure 1.

Insert Figure 1 about here

To avoid arbitrarily assigning subjects with GPA's at or near 3.0 to the low or high group, the low GPA group in this analysis was the lower one fourth of the subjects, and the high GPA group was the upper one quarter.

Figure 1 shows that the low GPA subjects' learning curve (plotted against number of errors on the teaching materials) was a shallow U-shaped curve: this group apparently did best with very few or with very many errors. The opposite was true for the high GPA subjects: their curve was an inverted U-shaped curve -- they did best with an intermediate number of errors, and learned least when the material was either very easy or very difficult for them. Though this test for different quadratic slopes was not quite significant at the .05 level (p=.096 when this model was reduced to the next one), the trend is definite. This is an interesting suggestion of a non-linear aptitude-treatment interaction. Possibly if the range of errors on the teaching materials had been greater, the trend of differential learning for the two GPA groups would have been more statistically significant.

Next, the three different difficulty level groups were combined into a single constructed response group which was compared with the reading-only group in an attempt to test response mode effectiveness. At least two factors lend justification to the combining of these groups: 1) the important commonalities between the different sets of materials -- number of frames, format, required responses, overall functional equivalency, etc.; and 2) the lack of significant learning differences between the groups, as reported above. GPA median splits were made in the reading mode and constructed response groups for the purpose of analyzing for a GPA by response mode interaction. Table 3 gives the means, standard deviations, and source data for the 2X2 factorial design.

Insert Table 3 about here

Again, we found a significant main effect for GPA, with the high GPA subjects doing better than the low. There also was a significant main effect for response mode, with the reading-only group scoring higher than the constructed response group (F=6.13, df=1,143, p=.01). No response mode by GPA interaction was found.

In reference to questions posed earlier, we may now infer that, in addition to the general treatment effect found, both response mode and number of errors on teaching materials are important as more specific issues. Also, high GPA subjects consistently learned more than the low GPA subjects. The GPA X treatment group interactions were not significant, but evidence was found for a non-linear GPA X error rate interaction.

The difficulty level groups did not differ significantly in learning as had been expected. One explanation would be that while difficulty is important, the range sampled here (3.75%, 18.75%, and 35.25% errors) was too restricted to produce effects. If the range of difficulty level were increased and a significant treatment effect appeared, one might hypothesize that different GPA groups would learn better with different types of materials: the low GPA subjects might do as well or better than the high GPA subjects with easy materials, but score lower when taught with difficult materials. This pattern would be expected given the Brophy and Evertson (in press) data, although extrapolation is difficult since their data were taken in teacher-controlled, naturalistic classroom environments, using subjects in the early concrete operational years.

Because of the "ceiling" effect present in the posttest data, methodological changes in future studies should also include the use of a more difficult posttest and/or less competent subjects.

Discussion

Without regard to the type of instructional materials used, the high GPA subjects scored higher than the low GPA subjects on the multiple choice posttest. This fits the Williams (1963) findings that high ability subjects learn more than low in experimental situations.

The evidence that low and high GPA groups have different non-linear patterns of learning plotted against number of errors on teaching materials is also of note. Achievement motivation theory, as stated by Atkinson and Feather (1966) and investigated by Koehler (Note 1) and others, would predict that subjects high in achievment motivation should do best when the instructional materials are of moderate difficulty. Conversely, low achievement motivation subjects

do best with (and tend to approach) tasks which are either very easy or very difficult. Apparently, low achievement motivation subjects experience anxiety which impairs the learning process when the difficulty level is moderate; when the task is very easy or very difficult, this anxiety is minimal (one is either likely to succeed -- easy task, or to escape culpability for failure -- very difficult task). High achievement motivation subjects perform best when the task is neither too easy (and probably boring) nor too difficult and frustrating.

When the subjects were grouped by GPA level, the near-fit of the data to a "different quadratic slopes" model is consonant with achievement motivation theory predictions. It seems that college GPA acts as a motivational measure: the high GPA subjects perform as a high achievement motivation group should perform, and the low GPA subjects' learning curve (plotted against number of errors on the teaching materials) is similar to the curve which would be predicted for low achievement motivation subjects. The data presented here also support the notion of an optimal difficulty level which may vary with the group or groups of learners under consideration; this is in line with the Brophy and Evertson (in press) findings.

Reading mode was found to be superior to constructed response mode in this study. At first, this result seemed counterintuitive and contradictory to previous findings. Many studies had indicated that the best that reading-only materials could do would be to equal the constructed response materials in producing learning. However, comparisons conducted by this author using the Tobias (1969) data and the Abramson and Kagen (1975) data, indicate that this is not the case in some situations. Both the Tobias (1969) and the Abramson and Kagen (1975) studies looked at familiarity of the subject matter

as an independent variable. Their findings of reading mode being better than constructed respone always were for familiar material.

If one considers the pretest given to subjects in this study to be a "familiarizing task," the superiority of the reading-only group makes sense. In fact, it could be argued that having subjects take a multiple choice pretest is more familiarizing than having them memorize a list of related technical terms, as Abramson and Kagen (1975) did. Thus, we see that the familiarity interpretation -- that constructed response is better for unfamiliar material, but reading mode is as good or better for familiar material -- is supported across three different studies, using different subjects and different teaching materials. Implications of this finding bear on instructional design: when the subject matter is new and unfamiliar, requiring the learner to construct responses is apparently most effective; after the topic has been introduced, when the learner is familiar with the material, having him or her read the material seems to allow for maximum learning.

Tobias (1973) mentioned "time on task" as possibly an important underlying factor in response mode effectiveness studies. If more time on the learning task always produces more learning, it would not seem possible for the reading mode ever to produce higher scores than constructed response (other things being equal, constructed response always takes longer than reading). If "time on task" is an operative factor, it should interact with type of material to be covered, familiarity, and sex of the learner (see Abramson and Kagen, 1975).

Another factor often discussed as relevant to this type of research is the "degree of active participation." The usual assumption is that,

the greater the degree of active participation, the greater the learning from the process of participation. The superiority of reading-only materials in some instances may question this assumption. Again, this participation factor probably is important, but it also probably interacts with learner characteristics and other instructional aspects of the environment.

One problem here seems to be that "learning by reading" usually is considered "passive" learning. The concept of "passive" learning becomes meaningless in any consideration of learning as an interactive process. It may be that having a subject write down a response is more cognitively active than reading that response with a line under it, but this is an assumption at present, not a documented fact. In any case, the optimal amount of "active participation" apparently is dependent on some of the other variables mentioned previously.

The most important general consideration here may be the matching of instruction to the learner's cognitive structures so as to optimize the learning process. That is, instructional variables (which are external to the learner) probably interact among themselves and and with factors which are internal to the learner. Response mode and difficulty level may not only interact with familiarity of the material and type of subject matter, but also with learner characteristics like ability, sex, cognitive development level, and with the learner's affective domain (e.g., fear of failure, achievement motivation). Consideration should be given to all these factors in designing effective instructional programs. The complexities inherent in this type of research are many, but the potential outcome -- a model for optimizing learning -- would seem to be worth continued efforts.

Reference Note

1. Koehler, V. The theory of achievement motivation, and grades and occupational aspirations. Unpublished doctoral dissertation, Syracuse University, 1973.

References

Abramson, T. & Kagen, E. Familiarization of content and different response modes in programmed instruction. Journal of Educational Psychology, 1975, 67, 83-88.

Atkinson, J. W. & Feather, N. T. A theory of achievement motivation. New York: John Wiley & Sons, 1966.

Brophy, J. & Evertson, C. Learning from teaching: A developmental perspective. Boston, Mass.: Allyn and Bacon, In press.

Crowder, N. The rationale of intrinsic programming. Journal of Programmed Instruction, 1962, 1, No. 5, 3-6.

Goldbeck, R. & Campbell, V. The effects of response mode and response difficulty on programmed instruction. Journal of Educational Psychology, 1962, 53, 110-118.

Holland, J. A quantitative measure for programmed instruction. American Educational Research Journal, 1967, 4, 87-101.

Holland, J. & Skinner, B. The analysis of behavior. New York: McGraw-Hill, 1961.

Morgan, C. Physiological psychology. New York: McGraw-Hill, 1965.

O'Day, E., Kulhavy, R., Anderson, W., & Malczynski, R. Programmed instruction techniques and trends. New York: Appleton Century Crofts, 1971.

Pressey, S. & Kinzer, J. Auto-elucidation without programming. Psychology in the Schools, 1964, 1, 359-365.

Pysh, F., Blank, S., & Lambert, R. The effects of step size, response mode, and knowledge of results upon achievement in PI. The Canadian Psychologist, 1969, 10, No. 1, 49-64.

Skinner, B. Teaching Machines. Scientific American, 1961, 205, 90-102.

Tobias, S. The effect of creativity, response mode, and subject matter familiarity on achievement from programmed instruction. Journal of Educational Psychology, 1969, 60, 453-460.

Tobias, S. Review of the response mode issue. Review of Educational Research, 1973, 43, 193-204.

Ward, J. & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

Williams, J. Comparison of several response modes in a review program. Journal of Educational Psychology, 1963, 54, 253-260.

Winer, B. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Footnotes

The author wishes to acknowledge and thank the following individuals who participated in the research described herein and/or in the preparation of this report: Dr. Jere E. Brophy, Dr. Carolyn M. Evertson, Dr. Jack Dunham, Cynthia Coulter, and Michael Baum, for their general consultation and comments; Gael Sherman and Linda Mahaffey Anderson, for their help as experimenters; Dr. Earl Jennings, for his consultation on the statistical analyses; Ann Turney, for her help in scoring the experimental packets; and Gwen Newman, for her assistance in manuscript preparation.

Table 1

Analysis of Covariance and Mean Scores

for the Four Treatment Groups

Mean Posttest Scores and Standard Deviations

		Reading	Low Error Rate	Medium Error Rate	High Error Rate
Transformed		N=39	N=39	N=45	N=36
Posttest	$\bar{x}$	1.405	1.399	1.374	1.363
Scores	SD	.073	.069	.058	.080

Treatment Main Effects Test

Model	R^2	SS (Error)	Mean Squares	F	p
Full	.2324	.6269	.0041	3.1671	.0262
Restricted	.1850	.6656			
Difference	.0474	.0387	.0129	df=3,154	

Table 2

Analyses of Covariance and Means

for the GPA by Difficulty Level Analyses

Mean Posttest Scores and Standard Deviations

			Low Error Rate	Medium Error Rate	High Error Rate
	Transformed		N=17	N=20	N=17
Low GPA	Posttest	$\bar{x}$	1.362	1.356	1.346
	Scores	SD	.079	.073	.051
	Transformed		N=18	N=20	N=17
High GPA	Posttest	$\bar{x}$	1.424	1.395	1.392
	Scores	SD	.041	.031	.068

GPA Main Effects Test

Model	R^2	SS (Error)	Mean Squares	F	p
Full	.3103	.3208	.0031	13.0916	.0005
Restricted	.2217	.3619			
Difference	.0885	.0412	.0412	df=1,102	

Difficulty Level Main Effects Test

Full	.3103	.3208	.0031	.9545	.3884
Restricted	.2973	.3268			
Difference	.0129	.0060	.0030	df=2,102	

GPA X Difficulty Level Interaction Test

Full	.3103	.3208	.0031	.0669	.9353
Restricted	.3093	.3212			
Difference	.0009	.0004	.0002	df=2,102	

Table 3

Analyses of Covariance and Means

for the GPA by Response Mode Analyses

Mean Posttest Scores and Standard Deviations

			Reading	Constructed Response
	Transformed		N=19	N=54
Low GPA	Posttest	$\bar{x}$	1.376	1.351
	Scores	SD	.085	.067
	Transformed		N=20	N=55
High GPA	Posttest	$\bar{x}$	1.433	1.407
	Scores	SD	.042	.050

GPA Main Effects Test

Model	R^2	SS (Error)	Mean Squares	F	p
Full	.2933	.4880	.0034	12.6469	.0005
Restricted	.2308	.5312			
Difference	.0625	.0432	.0432	df=1,143	

Response Mode Effects Test

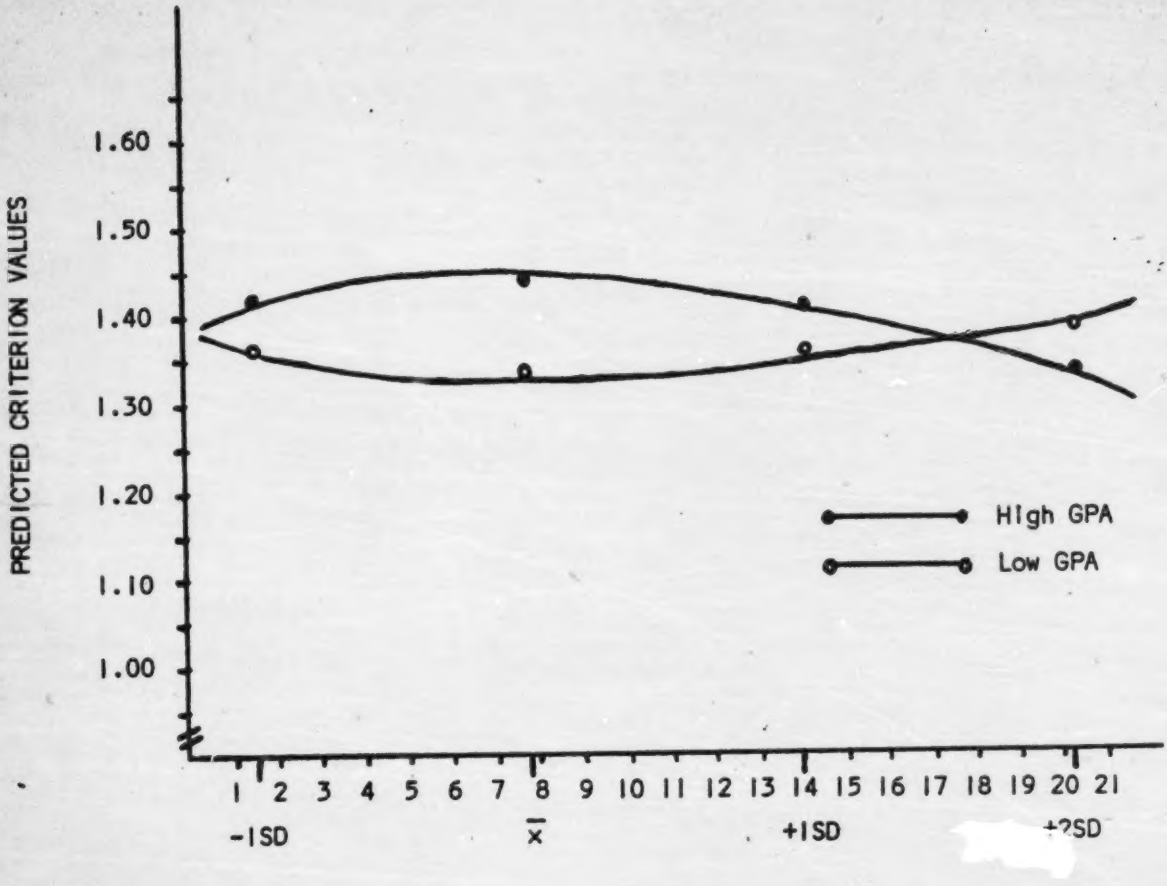
Full	.2933	.4880	.0034	6.1288	.0145
Restricted	.2630	.5089			
Difference	.0303	.0209	.0209	df=1,143	

GPA X Response Mode Interaction Test

Full	.2933	.4880	.0034	.3451	.5578
Restricted	.2916	.4892			
Difference	.0017	.0012	.0012	df=1,143	

Figure Caption

Figure 1. Regression lines of low and high GPA subjects plotted against number of errors on the teaching materials.


PREDICTED CRITERION VALUES
1.60
1.50
1.40
1.30
1.20
1.10
1.00
High GPA
Low GPA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
-1SD
$\bar{x}$
+1SD
+2SD
NUMBER OF ERRORS ON TEACHING MATERIALS